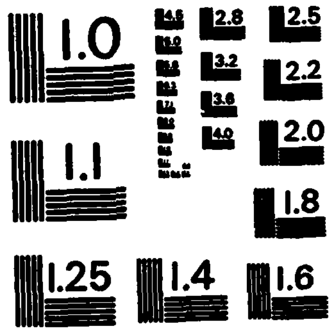
AD-A121 904 REPEATED MEASURES ON A CHOICE REACTION TIME TASK(U) 1/1
NAVAL BIODYNAMICS LAB NEW ORLEANS LA M KRAUSE ET AL.
AUG 82 NBDL-82R006

UNCLASSIFIED F/G 12/1 NL

END
FILMED
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

30

NBDL - NBDL-82R006

REPEATED MEASURES ON A CHOICE REACTION TIME TASK

M. KRAUSE AND A. C. BITTNER, JR.

AUGUST 1982

NAVAL BIODYNAMICS LABORATORY
New Orleans, Louisiana

Approved for public release. Distribution unlimited.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NBDL-82R006	2. GOVT ACCESSION NO. AD-A122 904 / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Repeated Measures on a Choice Reaction Time Task	5. TYPE OF REPORT & PERIOD COVERED Research Report
	6. PERFORMING ORG. REPORT NUMBER NBDL-82R006
7. AUTHOR(s) M. Krause and A. C. Bittner, Jr.	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Biodynamics Laboratory Box 29407 New Orleans, LA 70189	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS MF .58.524-002-5027
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Medical Research & Development Command Bethesda, MD 20014	12. REPORT DATE August 1982
	13. NUMBER OF PAGES 5
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Repeated Measures
Choice Reaction Time
Human Performance Testing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This study investigated a visual choice reaction time task to determine its suitability for repeated measures experimentation. Fifteen men served as subjects for one-, two-, and four-choice reaction time conditions. Fifty trials on each condition were administered for 15 days. Results indicated that across all RT conditions, intersession correlations were differentially stable by about Day 8, with reliabilities around .71. Means and variances remained unchanged over the stable days for the RT measure. It was concluded that studies using RT as the dependent variable should allow for at least 1000 practice

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Abstract

Block 20 (Continued)

trials prior to using RT as a performance assessment tool.

NBDL - 82R006

REPEATED MEASURES ON A CHOICE REACTION TIME TASK

M. KRAUSE AND A. C. BITTNER, JR.

AUGUST 1982

Bureau of Medicine and Surgery
Work Unit No. MF.58.524-002-5027

Approved by

Channing L. Ewing, M. D.
Chief Scientist

Released by

Captain L. E. Williams, MC USN
Commanding Officer

Naval Biodynamics Laboratory
Box 29407
New Orleans, LA 70189

SUMMARY PAGE

PROBLEM

Valid assessment of environmental effects on humans is dependent on stable measurement tools. Traditional measuring tools for the psychologist are reliable tests of cognitive and psychomotor functioning. Special statistical requirements must be met, however, before any of these tools are suitable to use for assessing before-, during-, and after-exposure environmental effects. Choice reaction time, a component of several Navy jobs, was examined to determine whether it met the statistical criteria necessary for repeated measures applications and thereby met the criteria for inclusion in a performance assessment battery.

FINDINGS

After approximately 1000 practice trials on visual choice reaction time, the task is suitable for repeated measures use.

RECOMMENDATIONS

Choice reaction time should be included in performance assessment batteries because it meets the necessary statistical criteria and additionally represents an ability that is vital to safety and defense.

This research work was funded by the Naval Medical Research and Development Command.

The volunteers used in this study were recruited, evaluated and employed in accordance with the procedures specified in the Secretary of the Navy Instruction 3900.39 series and the Bureau of Medicine and Surgery Instruction 3900.6 series. These instructions are based upon voluntary consent, and meet or exceed the prevailing national and international guidelines.

REPEATED MEASURES ON A CHOICE REACTION TIME TASK

Michele Krause and Alvah C. Bittner, Jr.
Naval Biodynamics Laboratory, New Orleans, Louisiana 70189

ABSTRACT

The current investigation examined a visual choice reaction time (CRT) task to determine its suitability for repeated measures experimentation. Fifteen Navy enlisted men served as subjects for one-, two-, and four-choice reaction time conditions. Fifty trials on each condition were administered for 15 consecutive workdays. Reaction time (RT), movement time (MT), and total time (TT) were recorded for each trial. Results indicated that across all RT conditions, intersession correlations were differentially stable by about Day 8, with reliabilities around .71. Means remained unchanged over the stable days and variances remained constant across all 15 days of the experiment for the RT measure. It was concluded that studies using RT as the dependent variable should allow for sufficient RT practice prior to data collection to ensure that the results are not contaminated by learning effects. It is recommended that at least 1000 practice trials be given prior to using an RT task to assess the effects of an environment.

INTRODUCTION

Reaction time (RT) has been a subject of study since Donders (1868) characterized the three basic types of tasks: simple, choice, and disjunctive. Since that time the literature has been abounding with theoretical propositions to account for the differences in RT performance across conditions (see Smith, 1968 for a review; Grice, Nullmeyer & Spiker, 1982). Equally as many studies have concentrated on identifying the components of RT tasks, and isolating factors which affect the subprocesses. Some well-documented variables affecting response latency are: stimulus probability (Hyman, 1953), stimulus-response compatability (Fitts & Dieninger, 1954), location probability (Niemi & Keskinen, 1980), and foreperiod and stimulus modality, duration, and intensity (Niemi & Naatanen, 1981). This body of evidence has been and continues to be an impetus for human factors considerations in equipment and environmental design. Given even the best human factors engineering, however, there are elements in an environment which may interact with the person to affect RT performance. Assessment of environmental effects on RT ability is of particular importance in military situations where the detection of a critical event and execution of an immediate response is vital. Since reaction time is a component of several real-world jobs, and therefore has face validity, it is a viable candidate for inclusion in an environmental assessment test battery. The present study focused on the effects of practice on RT performance. While the issue of practice is represented in the CRT literature, the studies have all been concerned with the convergence of response times across conditions. Teichner and Krebs (1974) presented a review of this literature and noted that after approximately one million trials, average RTs for one-choice through eight-choice conditions are essentially equivalent.

The current study extends the practice effect issue to discuss what happens to a single RT condition when it is repeatedly practiced. In this study, a visual CRT task was examined to determine when, if ever, the task became differentially stable. Differential stability, which is a necessary requirement for repeated measures analysis-of-variance, implies that a single factor is being measured and that it remains constant from one session to another. Differential stability is indicated by homogeneous intersession correlations. Additional assumptions for simple repeated-measures analysis require that group means remain constant or increase linearly over stable trials, with variances unchanging (Bittner & Carter, 1981).

The purpose of the present study was both to determine whether a visual CRT task represents a stable human ability, and whether it meets the statistical criteria necesary for inclusion in a test battery for repeated measurement.

METHOD

Subjects. The subjects were 15 Navy enlisted males meeting the health qualifications described by Thomas, Majewski, Ewing, and Gilbert (1978). All subjects were volunteers for environmental research experiments and were between the ages of 18 and 25.

Apparatus. An eight-choice reaction time device was constructed for this experiment from schematics provided by Teichner and Williams (1978)[1]. One portion of the device was a control panel which allowed for manually setting one of eight stimulus lights on the subject's response box to illuminate. In addition, it had a restart button for initiating new trials. Light emitting diodes (LEDs) displayed each resulting reaction time (RT) and total time (TT) to the experimenter.

The subject's response box, pictured in Figure 1, measured 22.0 cm by 17.0 cm. Eight red (LED) lights, 0.64 cm in diameter and 2.0 cm apart center-to-center, were placed horizontally across the box, slightly below the center. Lights were numbered one through eight, from right to left. Response keys (1.3 cm square) were positioned 3.0 cm below each light and were 2.0 cm apart center-to-center. One "home" key (also 1.3 cm square) was centered 2.5 cm below the row of response keys. Only the middle four stimulus lights and corresponding response keys (those numbered three, four, five, and six) were used during this experiment; the two on each extreme end were covered throughout. The response box was fixed to the table at which the subject sat.

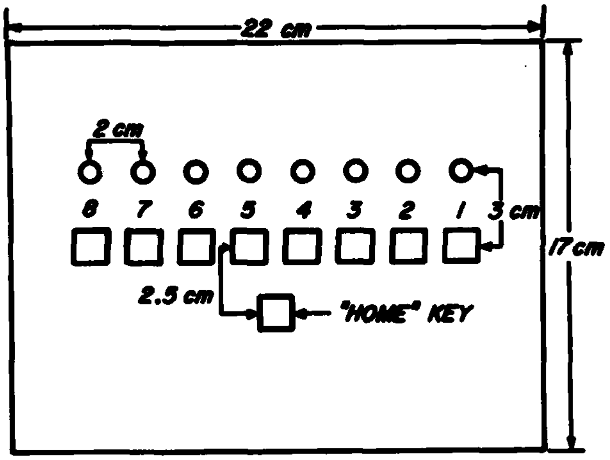


FIGURE 1: SUBJECT RESPONSE BOX

Procedure. Three days prior to experimentation, subjects were familiarized with the task. Instructions were read and the method was reviewed during this session, and again on the first testing day. Subjects were instructed to respond as quickly as possible without sacrificing accuracy. The task was individually administered in a quiet room, with the subject seated in front of the response box, perpendicular to the experimenter's control panel.

Fifty trials on each of three conditions: one-, two-, and four-choice, were presented in blocks. Stimulus light presentation was random and occurred approximately every ten seconds within a condition trial block. A 50 millisecond auditory warning signal followed by a 1500 millisecond delay preceded each stimulus presentation. The order of stimulus light conditions was randomized across days but remained the same for all subjects within a day. Subjects were tested for fifteen consecutive workdays. Therefore, over the course of the experiment, each subject was given 2250 trials, 750 at each condition. Two times were recorded: RT and TT. Reaction time (RT) was the time from stimulus onset to release of the "home" key. Total time (TT) was the time from stimulus onset to the time the button under the stimulus light was pushed. Movement time (MT) was calculated by subtracting RT from TT.

RESULTS

The analysis of results was conducted in four phases. The assumptions of the analysis were verified in the first phase. Secondly, the stability of the intersession correlations, means, and variances was examined. Correlations of the CRT conditions were compared and finally, subjects' response strategy was analyzed.

Verification of Analysis Assumptions

Autocorrelations were examined and the data was checked to determine whether a transformation was necessary. Autocorrelation patterns of selected subjects on random days indicated that the fifty successive RTs within a trial block were statistically independent; therefore, averages were used as the performance measure (the Box-Jenkins, 1970, approach was used). Ten successive scores within a trial block were averaged, reducing the data to five measures (replications) per condition per day.

An examination of selected extreme and some random data points revealed a nonsignificant correlation ($r = .29$) between RT means and corresponding variances. The implication was that a transformation of the data was unnecessary. This preliminary evidence guided the remaining analysis.

Stability Analysis

Differential Stability. Cross-session correlations were examined to determine when they became differentially stable. An overview of the results for each condition and measure, listed in Table 1, indicated stability on (approximately) Days 8 - 15 for RT and Days 10 - 15 for the MT and TT measures. Averaged cross-

TABLE 1: DIFFERENTIAL STABILITY RESULTS

COND. AND MEAS.		STAB. DAYS	AVG. REL.	CHI-SQUARED TEST*
1	RT	8-15	0.70	$\chi^2(27)=34.56$, $p = .14$
	MT	10-15	0.86	$\chi^2(14)=19.37$, $p = .15$
	TT	9-15	0.83	$\chi^2(20)=22.74$, $p = .30$
2	RT	8-15	0.63	$\chi^2(27)=36.26$, $p = .10$
	MT	10-15	0.86	$\chi^2(14)=15.61$, $p = .33$
	TT	7-15	0.80	$\chi^2(35)=40.32$, $p = .25$
4	RT	11-15	0.87	$\chi^2(9) = 5.21$, $p = .81$
	MT	9-15	0.91	$\chi^2(20)=20.51$, $p = .42$
	TT	10-15	0.91	$\chi^2(14)=15.92$, $p = .31$

* As assessed by the Steiger (1980) method

session reliabilities, shown along the diagonals in Table 2, were .71 for RT, .87 for MT, and .85 for TT.

Stability of Means. A separate Days (D) x Conditions (C) x Replications (R) ANOVA was run on RT, MT, and TT. Reaction time, the measure of primary interest, was examined first. Daily RT group means blocked across Conditions and Replications for RT were unchanged over Days 8 - 15 ($F(7,98) = 0.71$, $p = .67$). The Conditions main effect was highly significant, however, dominating all significant effects for the RT measure, with 95.9% of the total sum of the significant (C, R, CxR, DxCxR) sums-of-squares. As pictured in Figure 2, this effect was dominated

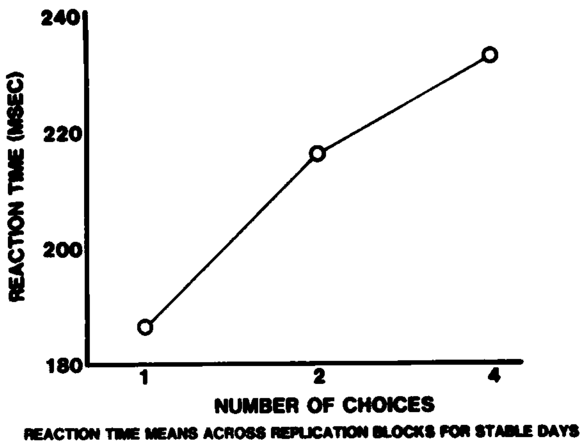


FIGURE 2: REACTION TIME MEANS ACROSS REPLICATION BLOCKS FOR STABLE DAYS

by a linear component ($F(1,14) = 51.64$, $p < .001$) which explained 97.5% of the variance. The quadratic component was also significant ($F(1,14) = 13.98$, $p = .002$) but insubstantial (2.5%). The overall Replications main effect was significant ($F(4,56) = 2.65$, $p = .04$). It was dominated by the linear component which accounted for 82% of the Replications variation. Relative to the Conditions effect, however, the Replications main effect accounted for less than .5% of the total sums-of-squares. The CxR interaction was significant ($F(8,112) = 3.33$, $p = .002$) but its explained sum-of-squares was less than 1% of the total. Figure 3 displays this interaction and reveals that differences between four- and two-choice conditions tended to shrink over replications while differences between one- and two-choices remained relatively constant. The DxCxR interaction was also significant ($F(56,784) = 1.60$, $p = .004$), and again only accounted for a fraction of the total significant sums-of-squares (2.8%). A second ANOVA revealed the nature of the MT measure. Daily group means across Conditions and Replications remained constant over Days 10 - 15 ($F(5,70) = .23$, $p = .95$).

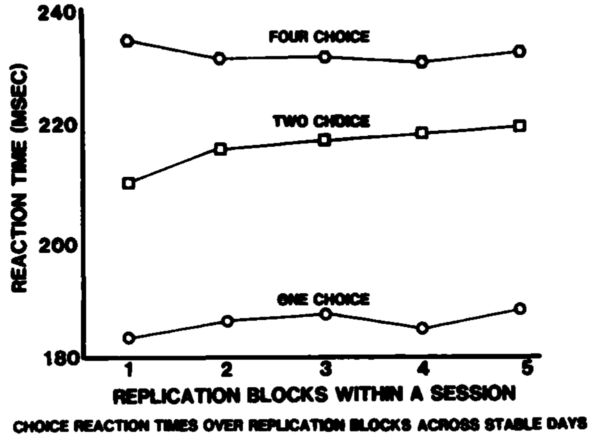


FIGURE 3: CHOICE REACTION TIMES OVER REPLICATION BLOCKS ACROSS STABLE DAYS

The Conditions main effect was also significant for the MT measure ($F(2,28) = 27.39$, $p < .0001$), with the linear component explaining 91.6% of the variance. Group means across stable days (10-15) for MT were 102.55 msec., 107.50 msec., and 120.05 msec. for the one-, two-, and four-choice conditions, respectively. All remaining simple and main effects for MT were nonsignificant. The final ANOVA indicated that TT daily group means across Conditions and Replications were stable across the last six experimental days ($F(5,70) = .91$, $p = .48$). Significance of the Conditions main effect carried through to the TT measure ($F(2,28) = 95.85$, $p < .001$). Almost 100% of the effect was due to linearity ($F(1,14) = 104.73$, $p < .001$). Overall, the Conditions effect accounted for 98.6% of the total significant sums-of-squares. Group means for TT were 287.31 msec. for one-choice, 323.35 msec. for two-choice, and 356.74 msec. for the four-choice condition. A significant CxR interaction ($F(8,112) = 3.45$, $p = .001$) accounted for the remaining insubstantial amount (1.4%) of the total variance. Fifty-four percent of the CxR interaction was accounted for by a linear-by-linear component ($F(1,14) = 9.66$, $p = .008$). Another 35% of the variance was attributed to a quadratic-by-quadratic component ($F(1,14) = 14.22$, $p = .002$). There were no other significant effects for the TT measure. In summary, the second phase of the analysis revealed that group means remained constant across stable days for RT, MT, and TT. Reaction time means stayed essentially the same after 7 days of practice, while MT and TT means ceased to change after Day 9. The Conditions main effect was highly significant for all measures, indicating that both processing and movement increased with the number of stimuli. The remaining significant effects accounted for an insubstantial amount of the total significant sums-of-squares.

Stability of Variances. Jackknife variance estimates (Carter & Bittner, 1982) were analyzed with separate Days (D) x Conditions (C) ANOVAs for each measure in the third phase of the analysis. Reaction time variances were stable over the course of the experiment ($F(14,196) = .37$, $p = .98$). There was no interpretable trend in the ANOVA for MT jackknife variances across the 15 experimental days ($F(28,392) = 1.58$, $p = .03$). The sources of significance were first-by-sixth, first-by-eighth, and second-by-second order interactions ($F(1,14) = 4.93$, $p = .04$, $F(1,14) = 16.67$, $p = .001$, and $F(1,14) = 13.09$, $p = .003$, respectively). The DxC interaction persisted, even when the four earliest days of practice were dropped ($F(20,280) = 1.66$, $p = .04$). Total time variances were stable across the 15 days ($F(14,196) = 1.11$, $p = .35$).

Comparisons Among CRT Conditions

Correlations between the three conditions (one-, two-, and four-choice) were examined over Days 10 - 14 for each measure (RT, TT, and MT). Table 2 indicates moderate to low correlations among RT conditions for the RT score, ranging from .25 to .65. In contrast, there was a high correlation (>.84) between conditions for MT. Total time correlations ranged from .74 to .83.

Within-day comparisons were separated from the average cross-day correlations, and appear in parentheses in Table 2. Within-day comparisons artificially inflate cross-day reliabilities by contributing variance that is unique to a day rather than unique to a condition. For two of the three cases, within-day comparisons significantly increased the cross-day correlations. For RT, MT, and TT the chi-squared values for the differences were $\chi^2(3) = 45.03$, $p < 10^{-9}$, $\chi^2(3) = 32.64$, $p < 10^{-6}$, and $\chi^2(3) = 3.73$, $p = .29$, respectively.

TABLE 2: CORRELATIONS ACROSS CONDITIONS

		1	2	4
RT	1	.78*	.65(.72)	.25(.33)
RT	2	.89**	.68*	.45(.62)
RT	4	.34**	.67**	.68*
MT	1	.91*	.87(.86)	.86(.89)
MT	2	.99**	.84*	.85(.91)
MT	4	.98**	.99**	.86*
TT	1	.91*	.83(.84)	.74(.79)
TT	2	.97**	.81*	.77(.86)
TT	4	.86**	.95**	.82*

* Cross-session reliabilities
** Corrected for attenuation correlations
Within-in day comparisons in parentheses

Examination of Strategy

Unusually low random RTs were observed throughout the data, which raised the question of whether subjects were lifting off the "home" key before they had perceived the stimulus light. If they were releasing the "home" key too soon, thereby perceiving the stimulus and initiating a response after they lifted off the "home" key, an unusually short RT and a relatively long MT would result. In the final phase of the analysis, a negative correlation between RT and MT was observed when (3) random subjects' data on (3) random days was examined in order to detect use of an early lift-off strategy. Figure 4 shows that this strategy was most apparent on the one- and two-choice conditions and was more evident after practice. It is intuitively reasonable that there would be more anticipatory responses, thereby more "false starts", when there were fewer stimulus choices and after the task had been learned and had become rather repetitious.

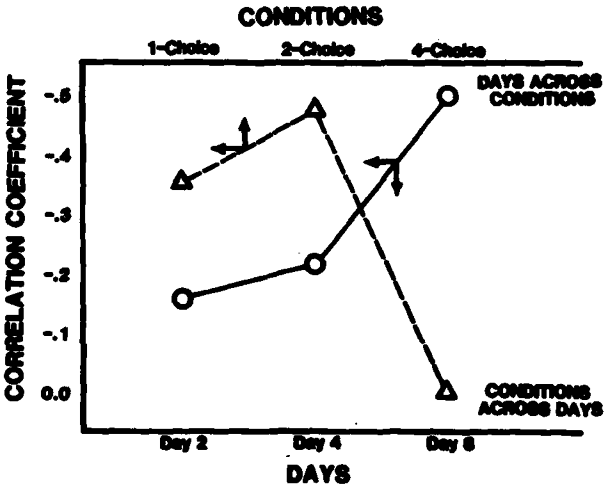


FIGURE 4: CORRELATION OF MT AND RT ON SELECTED DAYS

DISCUSSION

The following will address CRT stability, its usefulness as a performance assessment tool, and discuss general implications of the present research.

Stability. The results indicate that the RT and TT measures across all three conditions reached mean, variance, and differential stability during the course of the experiment. Means were level across Days 8 - 15 and Days 10 - 15, respectively, for all RT and TT conditions. Variances for both measures were stable across all Days (1 - 15). Differential stability was present by about Day 8 for RT and Day 10 for TT conditions. This implies that strategy changes ceased subsequent to Day 8. Altogether, across the three conditions, the RT and TT measures met mean, variance, and differential stability by about Day 8 and 10, respectively.

Mean and differential stability were evident by Day 10 for the MT measure. The variances for all MT choice conditions, however, were not constant across the 15 experimental days. The observed correlation between MT and RT suggests a possible source of this variation (c.f., Figure 4). Undetected sources of variance may also be operating. Movement time is an unstable measure because it lacks variance homogeneity.

CRT as an Assessment Tool. The RT and TT measures both appear suitable for inclusion as performance assessment tools, given sufficient practice. Mean, variance, and differential stability of these measures implies acceptability for repeated-measures applications (c.f., Bittner & Carter, 1981; Jones, Kennedy, & Bittner, 1981). An examination of the correlations among conditions (see Table 2) suggests the use of only one- and four-choice RT and TT conditions because of their measure of relatively unique variance. The TT measure is not recommended, however, because it is contaminated by the MT measure. For this reason, total time will be eliminated from further consideration in this report.

In the present experiment it was found that about seven days (1050 trials) of practice were required for RT to reach statistical suitability. It should be noted that the required practice could potentially be reduced by methodological changes. In particular, the strategy trade-off between MT and RT might be reduced or eliminated by using a random foreperiod. Additionally, an RT device which required the hands to be pre-positioned over the response alternatives would eliminate this problem. Appropriate feedback throughout might also be expected to reduce strategy shifts. Any methodological changes, however, would necessitate confirmation of stability characteristics. Overall, the RT measure for one- and four-choice conditions can be recommended as a tool in a performance test battery after sufficient practice.

Conclusion. The one- and four-choice RT measures are generally stable and are recommended for inclusion in performance assessment batteries, with at least 1000 practice trials prior to repeated-measures applications.

ACKNOWLEDGEMENTS

The author wishes to thank R. C. Carter for his contributions and support throughout this project. R. S. Kennedy's contributions to this research effort are also recognized. He initiated the study and saw it through the data acquisition and initial analysis stages before he left NBDL for Canyon Research Group, Orlando, FL. Thanks also to Lawrence Bell for aiding in data collection, E. Molino and his associates at NAMRL, Pensacola, G. Willems, and D. Knouse of NBDL for building the testing apparatus. Lastly, thanks to J. Woldstad for preparing the graphs and J. Marshall for typing the report.

NOTES

[1] A feasibility study run prior to this experiment indicated a significant inner vs. outer lights effect for the eight-choice but not the four-choice task. This experiment included only four- and fewer choices to avoid extraneous variation due to stimulus light position.

REFERENCES

Bittner, A. C., Jr. & Carter, R. C. Repeated measures of human performance: A bag of research tools. (Research Report No. NBDL-81 R011). New Orleans: Naval Biodynamics Laboratory. Nov. 1981. (NTIS No. AD A113954)

Box, G. E. P., & Jenkins, G. M. Time series analysis forecasting and control. San Francisco: Halden-Day, 1970.

Carter, R. C., & Bittner, A. C., Jr. Jackknife variance analysis of multifactor experiments. (Research Report No. NBDL-82R013) New Orleans: Naval Biodynamics Laboratory, May 1982.

Donders, F. C. Die Schnelligkeit Psychischer Processe. Archiv fur Anatomie und Physiologie und Vissenschaftliche Medizin, 1868, 657-681.

Fitts, P. M., & Dieninger, R. L. S-R compatability: Correspondence among paired elements within stimulus and response codes. Journal of Experimental Psychology, 1954, 48, 483-492.

Grice, G. R., Nullmeyer, R., & Spiker, V. A. Human reaction time: Toward a general theory. Journal of Experimental Psychology: General, 1982, 111, 135-153.

Hyman, R. Stimulus information as a determinant of reaction time. Journal of Experimental Psychology, 1953, 45, 188-196.

Jones, M. B., Kennedy, R. S., & Bittner, A. C. Jr. A video game for performance testing. American Journal of Psychology, 1981, 94, 143-152.

Niemi, P., & Keskinen, E. Visual stimulus intensity and location probability: Interactive effects on choice reaction time. Scandinavian Journal of Psychology, 1980, 21, 175-184.

Niemi, P., & Naatanen. R. Foreperiod and simple reaction time. Psychological Bulletin, 1981, 89, 133-162.

Steiger, J. H. Tests for comparing elements of a correlation matrix. Psychological Bulletin, 1980, 87, 245-251.

Teichner, W. H., & Krebs, M. J. Laws of visual choice reaction time. Psychological Review, 1974, 81, 75-98.

Teichner, W. H., & Williams, E. Personal Communication, 1978.

Thomas, D. J., Majewski, P. L., Ewing, C. L., & Gilbert, N. S. Medical qualification procedures for hazardous-duty aeromedical research. London: AGARD, 1977.

END

FILMED

1-83

DTIC